

CLAIMS

We claim:

1. A method for enabling a speech-based, Internet search comprising:

generating cohorts;

generating a language;

generating a merged grammar from the cohorts and language.
2. The method as in claim 1 wherein the language comprises an N-gram language.
3. The method as in claim 1 wherein the language comprises a finite state language.
4. The method as in claim 1 wherein the language comprises a context free grammar.
5. The method as in claim 1 wherein the language comprises a context sensitive language.
6. The method as in claim 1 wherein the grammar comprises a phoneme grammar.
7. The method as in claim 1 wherein the grammar comprises a word grammar.
8. The method as in claim 1 wherein generating the cohorts and language further comprises eliminating cohorts and parts of the language which are associated with web sites having low traffic flow.
9. The method as in claim 1 further comprising deleting repetitive cohorts or repetitive parts of the language.
10. The method as in claim 1 further comprising generating the cohorts and language from phonemes.
11. The method as in claim 1 further comprising:

comparing a phoneme associated with a spoken word against the merged grammar.
12. The method as in claim 11 wherein the merged grammar comprises one or more synonyms.
13. The method as in claim 11 wherein the merged grammar comprises one or more conjugates.
14. The method as in claim 11 wherein the merged grammar comprises part of speech identifiers.

15. The method as in claim 11 further comprising generating information correlated to the spoken word.
16. The method as in claim 15 wherein the information comprises a web site address.
17. A method for enabling a speech-based, Internet search comprising comparing a phoneme associated with a spoken word against a merged grammar.
18. The method as in claim 17 wherein the grammar comprises one or more synonyms.
19. The method as in claim 17 wherein the grammar comprises one or more conjugates.
20. The method as in claim 17 wherein the grammar comprises part of speech identifiers.
21. The method as in claim 17 further comprising generating information correlated to the spoken word.
22. The method as in claim 21 wherein the information comprises a web site address.
23. A system for enabling a speech-based, Internet search comprising:
- a cohort generator adapted to generate cohorts;
 - a language generator adapted to generate a language; and
 - a merging unit adapted to generate a merged grammar from the cohorts and language.
24. The system as in claim 23 wherein the language comprises an N-gram language.
25. The system as in claim 23 wherein the language comprises a finite state language.
26. The system as in claim 23 wherein the language comprises a context free language.
27. The system as in claim 23 wherein the language comprises a context sensitive language.
28. The system as in claim 23 wherein the grammar comprises a phoneme grammar.
29. The system as in claim 23 wherein the grammar comprises a word grammar.
30. The system as in claim 23 further comprising a statistical unit adapted to eliminate cohorts and parts of the language associated with web sites having low traffic flow.
31. The system as in claim 23 further comprising an optimization unit adapted to delete repetitive cohorts or repetitive parts of the language.
32. The system as in claim 23 further comprising:
- a recognition unit adapted to compare a phoneme associated with a spoken word against

the merged grammar.

33. The system as in claim 32 wherein the grammar comprises one or more synonyms.
34. The system as in claim 32 wherein the grammar comprises one or more conjugates.
35. The system as in claim 32 wherein the grammar comprises part of speech identifiers.
36. The system as in claim 32 wherein the recognition unit is further adapted to generate information correlated to the spoken word.
37. The system as in claim 36 wherein the information comprises a web site address.
38. A system for enabling a speech based, Internet search comprising:
 - a recognition unit adapted to compare a phoneme associated with a spoken word against a merged grammar.
39. The system as in claim 38 wherein the grammar comprises one or more synonyms.
40. The system as in claim 38 wherein the grammar comprises one or more conjugates.
41. The system as in claim 38 wherein the grammar comprises part of speech identifiers.
42. The system as in claim 38 wherein the recognition unit is further adapted to generate information correlated to the spoken word.
43. The system as in claim 42 wherein the information comprises a web site address.
44. The system as in claim 23 further comprising a grammar database, the database comprising:
 - one or more words or phonemes;
 - a probability value associated with each of the one or more of the words or phonemes; and
 - a pointer associated with each of the one or more words or phonemes.
45. The system as in claim 44 wherein the database further comprises distance values associated with one or more words or phonemes.
46. The system as in claim 44 wherein the database further comprises one or more web site indices associated with each pointer.
47. The system as in claim 46 wherein the database further comprises one or more web site addresses associated with each index.

48. The system as in claim 47 wherein the database further comprises usage weights, wherein each weight is associated with a web site address.
49. The system as in claim 44 wherein the database further comprises one or more synonyms of a word or phoneme.
50. The system as in claim 44 wherein the databases further comprises one or more conjugates of a word or phoneme.
51. The system as in claim 44 wherein the database further comprises one or more part of speech identifiers associated with one or more of the words or phonemes.
52. A programmed medium adapted to:
- generate cohorts;
 - generate a language; and
 - generate a merged grammar from the cohorts and language.
53. The programmed medium as in claim 52 wherein the language comprises an N-gram language.
54. The programmed medium as in claim 52 wherein the language comprises a finite state language.
55. The programmed medium as in claim 52 wherein the language comprises a context free language.
56. The programmed medium as in claim 52 wherein the language comprises a context sensitive language.
57. The programmed medium as in claim 52 further adapted to eliminate cohorts and parts of the language associated with web sites having low traffic flow.
58. The programmed medium as in claim 52 further adapted to delete repetitive cohorts or repetitive parts of the language.
59. The programmed medium as in claim 52 further adapted to compare a phoneme associated with a spoken word against the merged grammar.
60. The programmed medium as in claim 59 wherein the grammar comprises one or more synonyms.
61. The programmed medium as in claim 59 wherein the grammar comprises one or more conjugates.

62. The programmed medium as in claim 59 wherein the grammar comprises one or more part of speech identifiers.
63. The programmed medium as in claim 59 further adapted to generate information correlated to the spoken word.
64. The programmed medium as in claim 63 wherein the information comprises a web site address.
65. The programmed medium as in claim 52 wherein the medium comprises a magnetic storage medium.
66. The programmed medium as in claim 52 wherein the medium comprises a CD.
67. The programmed medium as in claim 52 wherein the medium comprises a digital storage device.
68. A programmed medium adapted to:

compare a phoneme associated with a spoken word against a merged grammar.
69. The programmed medium as in claim 68 wherein the grammar comprises one or more synonyms.
70. The programmed medium as in claim 68 wherein the grammar comprises one or more conjugates.
71. The programmed medium as in claim 68 wherein the grammar comprises part of speech identifiers.
72. The programmed medium as in claim 68 further adapted to generate information correlated to the spoken word.
73. The programmed medium as in claim 72 wherein the information comprises a web site address.
74. A grammar database comprising:

one or more words or phonemes;

a probability value associated with each of the one or more of the words or phonemes;
and

a pointer associated with each of the one or more words or phonemes.
75. The database as in claim 74 further comprising distance values associated with one or more words or phonemes.

76. The database as in claim 74 further comprising one or more web site indices associated with each pointer.
77. The database as in claim 76 further comprising one or more web site addresses associated with each index.
78. The database as in claim 77 further comprising usage weights, wherein each weight is associated with a web site address.
79. The database as in claim 74 further comprising one or more synonyms of a word or phoneme.
80. The database as in claim 74 further comprising one or more conjugates of a word or phoneme.
81. The database as in claim 74 further comprising one or more part of speech identifiers associated with one or more of the words or phonemes.
82. A database comprising one or more web site indices associated with one or more pointers.
83. A database comprising one or more web site addresses associated with one or more web site indices.
84. The database as in claim 83 further comprising usage weights, wherein each weight is associated with a web site address.
85. A database comprising synonyms of words or phonemes.
86. The database as in claim 85 further comprising conjugates of the words or phonemes.